AI1013- Programming for AI

Ronit Ranjan
ai24btech11028@iith.ac.in

1 Partial derivative of the error function with respect to $W_{k,1}^{(2)}$

Partial derivative of the error function with respect to $W_{k,1}^{(2)}$ is given by:

$$\frac{\partial \mathcal{L}}{\partial W_{k,1}^{(2)}} = \sum_{i=1}^{N} \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \times \frac{\partial \hat{y}_i}{\partial O_i} \times \frac{\partial O_i}{\partial W_{k,1}^{(2)}}$$

We will calculate each one of them

1. $\frac{\partial \mathcal{L}}{\partial \hat{\mathbf{y}}_\mathbf{i}}$:

$$\mathcal{L} = \frac{1}{2N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

$$\frac{\partial \mathcal{L}}{\partial \hat{y}_i} = \frac{1}{N} (\hat{y}_i - y_i)$$

2. $\frac{\partial \hat{y}_i}{\partial O_i}$:

$$\hat{y}_i = \sigma(O_i)$$

$$\frac{\partial \hat{y}_i}{\partial O_i} = \hat{y}_i(1 - \hat{y}i)$$

3. $\frac{\partial O_i}{\partial W_{k,1}^{(2)}}$:

$$O = ZW^{(2)}$$

$$\frac{\partial O_i}{\partial W_{k,1}^{(2)}} = Z_k$$

Finally, $\frac{\partial \mathcal{L}}{\partial W_{k,1}^{(2)}}$

$$\frac{\partial \mathcal{L}}{\partial W_{k,1}^{(2)}} = \frac{1}{N} \sum (\hat{y}_i - y_i) \times \hat{y}_i(1 - \hat{y}i) \times Z_k$$

$$\frac{\partial \mathcal{L}}{\partial W_{k,1}^{(2)}} = Z^T.\delta_2$$

where $\delta^{(2)} = (\hat{y} - y) \odot \sigma'(\hat{y})$

2 Partial derivative of the error function with respect to $W_{k,l}^{(1)}$

Partial derivative of the error function with respect to $W_{k,j}^{(1)}$ is given by:

$$\frac{\partial \mathcal{L}}{\partial W_{k,j}^{(1)}} = \sum_{i=1}^{N} \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \times \frac{\partial \hat{y}_i}{\partial O_i} \times \frac{\partial O_i}{\partial Z_i} \times \frac{\partial Z_i}{\partial H_i} \times \frac{\partial H_i}{\partial W_{k,j}^{(1)}}$$

We will calculate each one of them:

1. $\frac{\partial \mathcal{L}}{\partial \hat{y}_i}$:

$$\mathcal{L} = \frac{1}{2N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$
$$\frac{\partial \mathcal{L}}{\partial \hat{y}_i} = \frac{1}{N}(\hat{y}_i - y_i)$$

2. $\frac{\partial \hat{y}_i}{\partial O_i}$:

$$\hat{y}_i = \sigma(O_i)$$
$$\frac{\partial \hat{y}_i}{\partial O_i} = \hat{y}_i(1 - \hat{y}_i)$$

3. $\frac{\partial O_i}{\partial Z_i}$:

$$O = ZW^{(2)}$$
$$\frac{\partial O_i}{\partial Z_i} = W^{(2)}$$

4. $\frac{\partial Z_i}{\partial H_i}$:

$$Z_i = \sigma(H_i)$$
$$\frac{\partial Z_i}{\partial H_i} = Z_i(1 - Z_i)$$

5. $\frac{\partial H_i}{\partial W_{k,j}^{(1)}}$:

$$H = XW^{(1)}$$
$$\frac{\partial H_i}{\partial W_{k,j}^{(1)}} = X_{k,j}$$

Finally, $\frac{\partial \mathcal{L}}{\partial W_{k,j}^{(1)}}$:

$$\frac{\partial \mathcal{L}}{\partial W_{k,j}^{(1)}} = \frac{1}{N} \sum (\hat{y}_i - y_i) \times \hat{y}_i(1 - \hat{y}_i) \times W^{(2)} \times Z_i(1 - Z_i) \times X_{k,j}$$

$$\frac{\partial \mathcal{L}}{\partial W^{(1)}} = X^T . \delta^{(1)}$$

where

$$\delta^{(1)} = (\delta^{(2)} W^{(2)}) \odot \sigma'(Z)$$